

Breakout Sessions on Data Availability

Ranked Ideas/Needs

Traces, Synthetics, Apps, Benchmarks, Performance

- A. Need well-known common clearinghouse**
 - A. SNIA, USENIX have interest/effort in this area
 - B. Is lanl.gov enough? Mirrors? Is there a real challenge here?
- B. Conquer legal issues**
 - A. Throwing out bids that don't allow disclosure of perf. Data
 - B. Reduce possibility of data used for marketing
- C. Size/Space/Support Funding issues**
- D. Common Formats, collection programs, obfuscation programs**
 - A. Extensible (e.g. fits MPI-IO and strace and new types of traces)
 - B. Use common tools (e.g. gzip/bzip2) and formats (e.g. SNIA XML) where relevant
 - C. Anonymity of donors (but then do we believe the data?)
 - D. Metadata describing traces in terms of application domain, nprocs, etc.
- E. Analysis tools**
 - A. System-wide tools (e.g. IPM from NERSC, profiler not trace tool)
 - B. Extending existing tools to apply to HEC I/O (check into SNIA tools)
- F. One benchmark to rule them all**
 - A. Isn't this just useful for newspaper articles?
 - B. Observation that there are two roles for benchmarks, research and system testing
- G. Set of benchmarks that exhibit certain categories of behavior (profiles)**
 - A. Applications are evolving their I/O systems

Summary

Traces, Synthetics, Apps, Benchmarks, Performance

Voters: 30 Total (9 - .gov, 12 - .edu, 9 - .com)

Votes:

.gov ranking

1. Need well-known common clearinghouse (37)
2. Analysis tools (27)
3. Set of benchmarks that exhibit certain categories of behavior (profiles) (27)

.edu ranking

1. Analysis tools (43)
2. Common Formats, collection programs, obfuscation programs (42)
3. Set of benchmarks that exhibit certain categories of behavior (profiles) (34)

.com ranking

1. Analysis tools (32)
2. Common Formats, collection programs, obfuscation programs (25)
3. Set of benchmarks that exhibit certain categories of behavior (profiles) (24)